DOCUMENT RESUME

ED 463 297 TM 033 724

AUTHOR Ediger, Marlow
TITLE Assessing State Mandated Tests.
PUB DATE 2002-00-00
NOTE 8p.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Alternative Assessment; Elementary Secondary Education; Evaluation Methods; Standardized Tests; *State Programs; *Test Use; *Testing Programs
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

State mandated tests are being implemented in the public schools, but states differ greatly in the complexity of their tests, making comparisons very difficult. States may have widely different definitions of what counts as proficient, and it is evident that state standards are set arbitrarily. It is also important to consider the relationship of state standards to the National Assessment of Educational Progress results as well as questions related to a potential national curriculum. Educators looking for alternative forms of student evaluation have suggested student portfolios, which might be used for state testing and measurement. Portfolios based on constructivist ideas provide information about student learning at an every day level. States that do depend on state mandated tests must be concerned with reliability and validity, and they must be sure to test meaningfully what students have had an opportunity to learn. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 297

Assessing State Mandated Tests

Marlow Ediger

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
M. Ediger
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

TM033724

2

BEST COPY AVAILABLE

ASSESSING STATE MANDATED TESTS

State mandated tests are being implemented rather rapidly in the public schools. The purpose of these tests is to

1. notice and report pupil achievement.
2. publish the results from these tests in the media.
3. make comparisons among school districts and schools within the involved state.
4. weed out low performing schools or give pupils in these schools a chance to transfer out, to a well performing school.
5. provide diagnostic tools from test results to the classroom teacher of their pupils tested.
6. provide teachers with a listing of state mandated objectives. The teacher may then align the local curriculum with the mandated objectives.
7. establish inservice education programs for local teachers so that they may be able to assist pupils to achieve the state mandated objectives.
8. develop within teachers the desire to have high expectations for pupil achievement.
9. have teachers become conscious and motivated to achieve necessary skills to teach pupils to attain high standards of excellence in learning.
10. help teachers to become conscious of the testing and measurement movement as a means of improving instruction (Ediger and Rao, 2000, Chapter Nine).

Comparing Differences Among Diverse State Standards

States differ much from each other as to the complexity level of their respective tests. Thus, the test results from one state's set of standards may be high as compared to a different state which has low pupil test results. Olson (Education Week, February 20, 2002) wrote:

In North Carolina, for instance, 84% of fourth graders scored at the proficient level on the state test, while only 28% scored at that level on NAEP (National Assessment of Educational Progress). In Wyoming, the proportion of 4th graders scoring at the proficient level on both the state and national level was closely matched, at 27% and 25% respectively.

Only idaho, Louisiana, Missouri, North Dakota, and Rhode Island had a smaller share of students scoring at the proficient level on their tests than on the NAEP at the fourth and eight grade.

That states may have widely different definitions of what counts as proficient has been pointed out since at least 1996. That's when Mark S. Musick, the president of the Atlanta based Southern Regional Education Board, wrote a report in which he noted that "state standards for student achievement are so dramatically different that they simply don't make sense"

Mr. Musick reached his conclusions after comparing the per cent of students who scored at the proficient level on state reading and mathematics in 1994- 1995 with the proportion who scored on the proficient level on NAEP. Only 13% of Delaware's 8th graders met the state's 8th grade math standard, compared with 83% of 8th graders in Georgia. Yet on the state NAEP, 8th graders in Delaware outscored Georgia counter parts. What's going on here he asked... "I have argued that state leaders should want to know why standards based results are so different. When they know why, then they can decide if they believe their standards are about right or whether they need to be changed."

From the above direct quote, it is quite obvious that state standards are set arbitrarily. Perhaps, this is true of all state standards for pupils to achieve. It is also true of the NAEP. Who is to decide which levels pupils should achieve in any academic discipline? In addition, the following questions are relevant to consider:

1. should state pupil test results be compared with other states in the union when there is much variance in results in comparison with NAEP?
2. should the difficulty level of each state's tests be reevaluated? This is crucial in high stakes testing whereby a pupil who fails may not receive a diploma for graduation.
3. should the level of difficulty of state mandated tests be more realistic? It is one thing to desire a certain level of pupil achievement whereas pupils are not ready to perform at that level of complexity.
4. should state objectives be more clearly written so that teachers may understand what might be covered in a mandated test?
5. should test items on each state's assessment be reevaluated in terms of validity and reliability? There might well be test items which do not cover what has been taught in a classroom. Thus, validity is lacking. The tests may not measure consistently; reliability is then lacking.
6. should each state mandated test be thoroughly pilot tested? In pilot studies, data may be obtained on test/ retest, alternate forms, and/or split half reliability.

7. should each state list the standard error of measurement for their tests? This is important in that the observer may then notice how much error in measurement there is on a state mandated test.

8. should more faith be placed on alternatives to testing to notice pupil achievement? A single test score is hardly enough evidence to ascertain how well pupils are achieving.

9. should state mandated tests be omitted and the NAEP take its place? Comparisons are made as to how state mandated tests differ from NAEP in terms of percent passing each when making state by state comparisons.

10. should a national curriculum be developed and implemented so that a nation wide mandated test may be given? This would tend to eliminate selected problems that exist when each state writes their very own tests (Ediger and Rao, 2001, Chapter Sixteen).

Thus, there are a plethora of questions which need answering pertaining to state mandated testing. These are not easy questions to answer. It appears that for every action taken in state mandated testing, there is an opposite and equal reaction.

Alternative Forms of Pupil Evaluation

Educators looking for alternative forms of pupil evaluation have identified a portfolio replacement. Portfolios might also be used in addition to state testing and measurement. The two approaches differ form each other in philosophy involved.

The testing and measurement movement emphasizes a philosophy of realism. Realists stress a scientific approach in dealing with knowledge. Thus, the observer can know the real world in whole or in part, as it truly is. For example, chemists have identified 106- 107 elements making up the planet earth. Elements can be combined to form molecules. Thus, for example, the formula for sugar is C6 H12 O6. Six atoms of carbon, 12 atoms of hydrogen, and six atoms of oxygen is the formula for a molecule of sugar. Exactness and precision are then inherent in measurement. The behaviorally state objectives movement has it basis in realism in that

1. each objective for pupils to achieve needs to be stated with precision.

2. the learning opportunities must be aligned for pupils to achieve these objectives.

3. measurement and testing to ascertain if these objectives

have been achieved is necessary to notice what pupils have learned.

4. the objectives of instruction need to be arranged in ascending order of complexity. Careful sequencing is wanted.

5. a numerical score provides the exact answer as to where a pupil is achieving. The numeral may be a percentile or a percent (See Ediger, 2002, pp 20-21).

Statewide testing omits pupil achievement reports from the every day work in class which learners do. The teacher has no input into test items content, time limits in giving the test, pilot study involvement, and/or modifications of the test. Portfolios take care of selected problems involved here. A portfolio then emphasizes constructivism/existentialism tenets in that the

1. the pupil with teacher guidance selects products/processes which should go into the personal portfolio to indicate that which has been learned.

2. a random sampling of items are then chosen for the portfolio.

3. everyday classroom work is selected to be represented in portfolio content.

4. parents and other responsible individuals might then view portfolio items to notice pupil achievement and progress.

5. portfolios are to be assessed by professionals in the field of teaching and learning. The following are weaknesses in advocating portfolio use to appraise pupil achievement:

1. they are difficult to assess and cannot be machine scored. Since human evaluators need to assess the portfolio, much time is spent in the assessment process. If these are paid assessors, the expenses could be great, indeed, in the assessment program.

2. interrater reliability could be low. Thus, two or more assessors for the same portfolio may come up with quite different results in its scoring.

3. it is difficult with the many entries for an evaluator to notice which products and processes pertain to any single objective of instruction.

4. rubric use may cut down on some of the subjectivity in the assessment process. But, rubrics generally contain rather broad criteria to use in their evaluation by raters.

5. too many entries in a portfolio make for a time consuming assessment activity (Ediger, 1994, 31- 43).

Portfolio advocates need to view the above five named weaknesses and work in the direction of taking out kinks. Weaknesses identified in either the testing/measuring approach or in portfolio use, provide healthy suggestions in working toward overcoming these problem areas.

Suggestions for Developers of State Mandated Tests

Those in charge of developing state mandated tests should not become overly ambitious in establishing complex objectives for pupil attainment. The objectives should be challenging, but achievable. Each pupil needs to achieve as much as possible. What is desired by the state for pupils to achieve may not be possible in reality. Establishing state standards and objectives for pupils to attain is not a science, but an art. People choose which standards and objectives pupils are to achieve. They do the analyzing and writing. Truth is in the eye of the beholder.

What has been taught and learned meaningfully may be tested in order for validity to be present. The state, too, needs to be careful that adequate reliability is there when tests are adopted to evaluate pupil achievement. Thus, consistency of test results from any one pupil is important. From pilot studies, the standard error of measurement needs to be spelled out clearly by the state. If the standard error of measurement is large, then specific cut off points for high stakes testing should not be enforced. Basing state tests results and their goodness upon NAEP findings has its problems. One being that one of the two tests should then be omitted since the NEAP is used to judge the quality of the state mandated test results. That must mean that the state mandated test does not have the merit which NAEP has.

States need to test meaningfully what pupils have had opportunities to learn, as listed in their objectives of instruction and these must be available to all teachers to use as guidelines for teaching. To use a single test for all pupils in a state violates the concept of providing for individual differences. Pupils differ from each other in a plethora of ways, one being the intelligences possessed. Testing emphasizes the use of verbal intelligences as in reading. Others may excel in music, art, physical skills as in athletics and dance, among others (See Gardner, 1993). Pupils, too, differ from each other in abilities possessed. A single standard such as a state mandated test does not appear to provide or individual differences. The handicapped child may need accommodations such as having

more time to complete test items (See Searson and Dunn, 22- 26).

There are numerous problems which need identification and solutions pertaining to testing and measuring pupil achievement, be it state mandated or NAEP tests. High quality tests with desired validity and reliability data need to be in the offing.

References

Ediger, Marlow (1994), "Philosophy in Teacher Education Programs," The Journal of Teaching Practice, 14 (2), 31- 43.

Ediger, Marlow, and D. Bhaskara Rao (2001), Teaching Social Studies Successfully. New Delhi, India: Discovery Publishing House, Chapter Sixteen.

Ediger, Marlow, and D. Bhaskara Rao (2000), Teaching Mathematics Successfully. New Delhi, India: Discovery Publishing House, Chapter Nine.

Ediger, Marlow (2002), "Writing Achievement in Technical Education," ATEA Journal, 29 (3), 20- 21.

Gardner, Howard (1993), Multiple intelligences: Theory Into Practice. New York: Basic Books.

Olson, Lynn (February 20, 2002), "A Proficient Score Depends Upon Geography," Education Week, 21 (23), pp 1, 14, 15.

Searson, Robert, and Rita Dunn (2001), "The Learning Styles Teaching Model," Science and Children, 39 (5), 22- 26.

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC/REC
2805 E. Tenth Street
Smith Research Center, [redacted] 140
Indiana University
Bloomington, IN 47408

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)